AD-A102 759 PENNSYLVANIA STATE UNIV UNIVERSITY PARK DEPT OF STAT--ETC F/G 12/1
WEIGHTED LEAST SQUARES RANK ESTIMATES.(U)
JUL 81 K CHENG, T P HETTMANSPERGER
UNCLASSIFIED TR-38 NL

1 OF 1
AD
A102759

END
DATE
FILMED
9-81
DTIC

LEVEL II

12

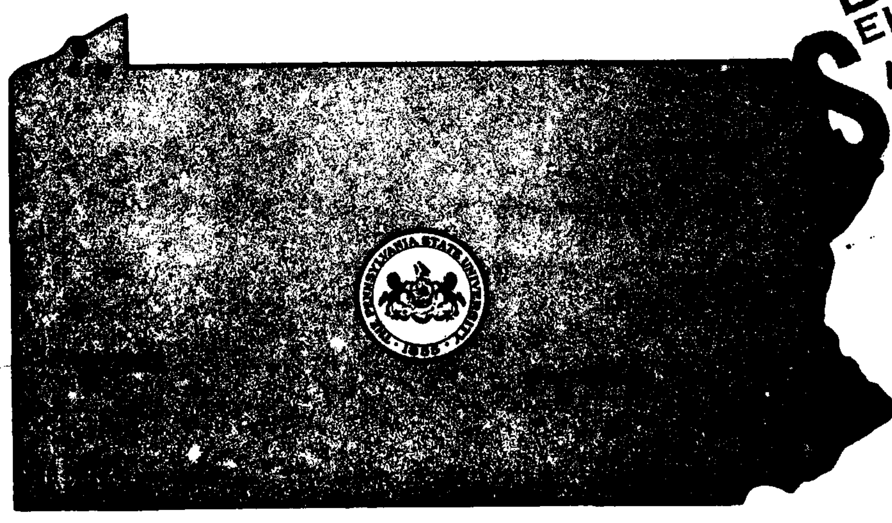
AD A102759

The Pennsylvania State University

Department of Statistics

University Park, Pennsylvania

DTIC FILE COPY



DEPARTMENT OF STATISTICS
The Pennsylvania State University
University Park, PA 16802, U.S.A.

12

TECHNICAL REPORTS AND PREPRINTS
Number 38: July 1981

WEIGHTED LEAST SQUARES RANK ESTIMATES

Kuo-Sheuon Cheng
Tulane University

Thomas P. Hettmansperger*
The Pennsylvania State University

*Research partially supported by ONR contract N00014-80-C-0741

Abstract

In this paper rank estimates, called WLS rank estimates and computed using iteratively reweighted least squares, are studied. They do not require the estimation of auxilliary scale or slope parameters nor do they require numerical search techniques to minimize a convex surface. The price is a small asymptotic efficiency loss. In the location model, beginning with a resistant starting value such as the median, the WLS rank estimates have good robustness and computational properties. The WLS rank estimate is also extended to the regression model and an example is given.

1. Introduction and Summary

To date there have been two major methods used to construct rank estimates in the linear model. The first is the direct minimization of the appropriate dispersion surface proposed by Jaeckel (1972). The minimization is equivalent to solving a set of non-linear equations and can be thought of as an extension of the method of Hodges and Lehmann (1963) for defining R-estimates of location. These estimates then have the same asymptotic efficiency properties as the rank estimates in the location case. In the location model, R-estimates generally satisfy various criteria for robustness such as bounded influence curves and positive breakdown values. See Hettmansperger and Utts (1977) and Huber (1981) for details of robust estimation.

The second method of construction consists in developing linearized versions of the original estimates of Jaeckel. This method uses the asymptotic linearity of rank test statistics developed by Jureckova (1971). Jureckova (1971) and Jaeckel (1972) used the linearity to derive the asymptotic distribution theory of the original rank estimates but did not use the linearity for the actual construction of estimates.

Kraft and van Eeden (1970), (1972a), (1972b), were the first to develop linearized rank estimates. Their estimates involve a starting value along with a scale estimate. In general these estimates are not quite as efficient as the nonlinear rank estimates. McKean and Hettmansperger (1978) develop linearized rank estimates for use in the linear model which have the same asymptotic efficiency as the nonlinear versions. These linearized estimates require the estimation of the slope of the linear approximation

Due to the time required to locate the minimum of the dispersion surface and determine the nonlinear rank estimate, it appears that some alternative is necessary. The practical value of the linearized rank estimates is pointed out in the references of Kraft and van Eeden (1972b) and McKean and Hettmansperger (1978). The question of how soon the asymptotic linearity provides an adequate approximation still remains. Further, the effect on efficiency in small samples of estimating scale or slope is not fully understood.

The M-estimate approach described by Huber (1981) provides another method of robust estimation. As in the rank estimation problem M-estimation requires the solution of nonlinear equations. Linearized versions of M-estimates have been studied by Bickel (1975). Andrews (1974) discussed some of the computational aspects of various approaches including weighted least squares. Interestingly he pointed out that iterating to convergence may be less desirable than simply using a fixed number of iterations from a good starting value. Finally Holland and Welsch (1977) have reported in detail on the iteratively reweighted least squares approach. They point out that the linearized version is theoretically more desirable than the weighted least squares; however, it is more difficult to implement because of the need to estimate the slope. M-estimation requires the estimation of a scale to make the resulting location and regression parameter estimates scale invariant. In the Holland-Welsch study scale was estimated just once with no further iterations because there is no convergence theory when scale is iterated along with the location estimates. Their Monte Carlo results indicate that for small samples estimation of scale has a strong

effect on the efficiency. As they point out, "In general, the effect of estimating the scale has been swept under the rug in previous studies of robust estimation and perhaps these results will bring attention to the fact that it should be more carefully considered." The situation seems to suggest that if a reasonably good estimate of scale can be incorporated into a weighted least squares M-estimate approach the result would be a computationally tractable estimator with fairly high efficiency.

In this study we study weighted least squares rank estimates, defined in Section 2. Unlike M-estimates, these estimates do not require the estimation of auxilliary scale functionals. They do not require the estimation of scale or of slope as in the case of linearized R-estimates nor do they require numerical search techniques to minimize a convex surface. In the location model, beginning with a resistant starting value such as the median, they have good efficiency and robustness properties; see Sections 3 and 4. In Section 5 we extend the procedure to the regression model and discuss an example.

2. The Asymptotic Distribution of WLS-Rank Estimates

For a given random sample $X_1, X_2, \ldots, X_n$ from a continuous symmetric distribution $G(x - \theta)$, where θ is unknown, an R-estimate of θ is a value which minimizes

$$S(\theta) = \sum_{i=1}^{n} a(R_i^+(\theta))|X_i - \theta| \tag{1}$$

where $0 \leq a(1) \leq \ldots \leq a(n)$ are constants, usually called scores and $R_i^+(\theta)$ is the rank of $|X_i - \theta|$ among $|X_1 - \theta|, \ldots, |X_n - \theta|$. Below we will show that the R-estimate can be considered as a weighted least squares estimate with weights proportional to the ranks of the absolute deviations.

First, we define, equivalently, an R-estimate $\hat{\theta}$ of θ as the solution of the following equation

$$h(\theta) = \sum_{i=1}^{n} a(R_i^+(\theta)) \operatorname{sign}(X_i - \theta) = 0 \tag{2}$$

Note that $h(\theta)$ is a nonincreasing step function of θ; see Bauer (1972). The R-estimate obtained is the Hodges-Lehmann (1963) estimate since $h(0)$ is a signed rank test statistic for testing H_0: $\theta = 0$ vs. H_1: $\theta > 0$. Except for special cases like $a(i) = 1$ with $\hat{\theta} = \operatorname{med} X_i$ or $a(i) = i$ with $\hat{\theta} = \operatorname{med} (X_i + X_j)/2$ solving the nonlinear equation (2) for θ is usually quite difficult. For example, for the van der Waerden or normal scores, there is no simple form for the R-estimate.

Hettmansperger and Utts (1977) in writing (2) as

$$\Sigma w_i(\theta)(X_i - \theta) = 0 \tag{3}$$

with

$$w_i(\theta) = \begin{cases} a(R_i^+(\theta))/|X_i - \theta| & \text{if } X_i \neq \theta \\ 0 & \text{otherwise} \end{cases}$$

were able to use an iteration procedure to obtain an R-estimate $\hat{\theta}_1$:

$$\hat{\theta}_1 = \frac{\Sigma w_i(\hat{\theta})X_i}{\Sigma w_i(\hat{\theta}_o)} = \hat{\theta}_o + \frac{h(\hat{\theta}_o)}{\Sigma w_i(\hat{\theta}_o)} \tag{4}$$

where $\hat{\theta}_o$ is an initial estimate of θ. Applying the same formula, the k-step estimate $\hat{\theta}_k$ can be obtained

$$\hat{\theta}_k = \hat{\theta}_{k-1} + \frac{h(\hat{\theta}_{k-1})}{\Sigma w_i(\hat{\theta}_{k-1})} . \tag{5}$$

We shall call them weighted least squares rank estimates or in short, WLS-rank estimates. To insure convergence, Utts (1978) proposed an algorithm which combined iteration with an interval halving procedure. She then proved the convergence of the k-step WLS-rank estimate to the nonlinear rank estimate.

In this paper, we discuss the efficiency and robustness properties of the WLS-rank estimates. Although similar to the Kraft and van Eeden estimate, in general the one-step WLS-rank estimates are not quite as efficient as the nonlinear rank estimates; the efficiency for the k-step estimate converges rapidly to that of the nonlinear estimate as k increases. Most of the WLS-rank estimates considered have bounded influence. Hence the weighted least squares rank estimation procedure provides a computationally feasible way to find robust R-estimates.

The proof of the asymptotic normality of the WLS-rank estimates will closely follow that of Kraft and van Eeden (1970).

Suppose we observe a sequence $X_1, \ldots, X_n$ of independent random variables such that $P_r(X_i \leq x) = G(x - \theta)$, $i = 1, \ldots, n$. Here the cumulative distribution function G is unknown but is assumed to be a member of the class Ω of distributions with absolutely continuous, symmetric densities with positive and finite Fisher information.

Let $F \in \Omega$ and define

$$\phi_f(u) = -\frac{f'}{f}(F^{-1}(u)) \qquad 0 < u < 1.$$

Assume that ϕ_f satisfies $\phi_f = \phi_1 + \phi_2$ where ϕ_1 is nondecreasing and ϕ_2 is nonincreasing, $\int_o^1 \phi_1^2 du < \infty$ and $\int_o^1 \phi_2^2 du < \infty$, and

$$\phi_f(1 - u) = -\phi_f(u) \tag{6}$$

Define $\phi_f^+(u) = \phi_f[(u + 1)/2]$ and consider a rank statistic $h_f(\theta)$ of the form

$$h_f(\theta) = \Sigma\phi_f^+\left(\frac{R_i^+(\theta)}{n + 1}\right) \text{sign } (X_i - \theta) \tag{7}$$

Let $\hat{\theta}_o$ be a consistent estimate of θ, then the one-step WLS-rank estimate is defined by

$$\hat{\theta}_1 = \hat{\theta}_o + \frac{h_f(\hat{\theta}_o)}{\Sigma w_i(\hat{\theta}_o)} \tag{8}$$

where
$$\omega_i(\theta) = \begin{cases} \dfrac{\phi_f^+\left(\frac{R_i^+(\theta)}{n + 1}\right)}{|X_i - \theta|} & |X_i - \theta| > 0 \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

Following Kraft and van Eeden (1970, 1972a) we will suppose for convenience, that the initial estimate $\hat{\theta}_o$ is asymptotically equivalent to a solution of another equation

$$h_S(\theta) = 0 \tag{10}$$

where $S \in \Omega$ is a fixed distribution function. Denote the Fisher information by $I(f) = \int\phi_f^2(u)du$, we then have

$$n^{-1/2}h_f(0) \xrightarrow{D} N(0, I(f)) \tag{11}$$

provided $I(F) < \infty$; see Hajek and Sidak (1967, p. 167).

Van Eeden (1972) obtained the asymptotic linearity for the signed rank statistic analogous to that of Jureckova's (1969) result. Using a special case of this result, Kraft and van Eeden (1970, 1972a) derived a linearized rank estimate for the center of symmetry in the one-sample location problem. We shall state a special case of Theorem 3.3 in the paper by van Eeden (1972) as a lemma.

<u>Lemma</u>. Suppose $G \varepsilon \Omega$ and ϕ_f satisfies the conditions above then

$$\lim_{n\to\infty} P\left(\sup_{|\theta-\theta_o| \leq cn^{-1/2}} n^{-1/2} \mid h_f(\theta) - h_f(\theta_o) + n(\theta - \theta_o)I(f(g) \mid > \varepsilon\right) = 0$$

for $\varepsilon > 0$, $c > 0$ a constant, and $I(f,g) = \int \phi_f(u)\phi_g(u)du$.

From this lemma we conclude that if $n^{1/2}\hat{\theta}$ is bounded in probability then

$$h_f(\hat{\theta}) = h_f(\theta) - n(\hat{\theta} - \theta)I(f,g) + o_p(n^{1/2}) \tag{12}$$

Notice that if we replace $h_f(\hat{\theta}) = 0$ by the linear approximation then we get $\hat{\theta} = \theta + \dfrac{h_f(\theta)}{nI(f,g)}$. Hence a one-step linearized rank estimate can be written as

$$\hat{\theta}_1 = \hat{\theta}_o + \frac{h_f(\hat{\theta}_o)}{n\hat{I}(f,g)} \tag{13}$$

where $\hat{I}(f,g)$ is an estimate of $I(f,g)$.

McKean and Hettmansperger (1978) estimated $I(f,g)$ directly and showed that there was no asymptotic efficiency loss relative to the nonlinear rank estimate for their estimates.

We now establish the asymptotic normality of the one-step WLS-rank estimate in the following theorem.

Theorem 1. If $G \varepsilon \Omega$, ϕ_f and ϕ_s satisfy the conditions above, $\hat{\theta}_o$ satisfies (10) and $n^{1/2}(\hat{\theta}_o - \theta)$ is bounded in probability, then

$$n^{1/2}(\hat{\theta}_1 - \theta) \xrightarrow{D} N(0,\ V_1(s,f,g)), \tag{14}$$

where

$$V_1(s,f,g) = \frac{I(S)}{I^2(s,g)} \left(1 - \frac{I(f,g)}{J(f,g)}\right)^2 + \frac{2I(s,f)}{I(f,g)I(s,g)} \left(1 - \frac{I(f,g)}{J(f,g)}\right) + \frac{I(f)}{I^2(f,g)} \tag{15}$$

and

$$J(f,g) = \int \frac{\phi_f(G(x))}{x}\, g(x)dx$$

provided $J(f,g)$ exists.

Proof. Without loss of generality, we assume that $\theta = 0$, then using the lemma,

$$\hat{\theta}_1 = \hat{\theta}_o + \frac{h_f(\hat{\theta}_o)}{\Sigma w_i(\hat{\theta}_o)}$$

$$= \hat{\theta}_o\left(1 - \frac{I(f,g)}{n^{-1}\Sigma w_i(\hat{\theta}_o)}\right) + \frac{n^{-1}(h_f(0) + o_p(n^{1/2}))}{n^{-1}\Sigma w_i(\hat{\theta}_o)}$$

using $R_i^+(\hat{\theta}_o) = nG_n^+(|X_i - \hat{\theta}_o|)$, and since $G_n^+(x) \xrightarrow{P} G^+(x)$, $\hat{\theta}_o \xrightarrow{P} 0$, $n^{-1}\Sigma w_i(\hat{\theta}_o) \xrightarrow{P} \int \frac{\phi_f^+(G^+(|x|))}{|x|}\, g(x)dx = J(f,g)$. Apply the asymptotic linearity to $h_S(\theta)$ to get

$$\hat{\theta}_o = \frac{h_S(0)}{nI(s,g)} + o_p(n^{-1/2}).$$

Hence

$$n^{1/2}\hat{\theta}_1 = \frac{1}{I(s,g)} \left(1 - \frac{I(f,g)}{J(f,g)}\right)n^{-1/2}h_S(0) + \frac{1}{J(f,g)}\, n^{-1/2}h_f(0) + o_p(1) \tag{16}$$

Following the asymptotic theory of rank tests (Hajek and Sidak, 1967, p. 166), we finally have

$$n^{1/2}\hat{\theta}_1 \xrightarrow{D} N(0,\ V_1(s,f,g)).$$

Similarly we can prove the following.

Theorem 2. Suppose $G \varepsilon \Omega$, ϕ_f and ϕ_s satisfy the above conditions, $\hat{\theta}_o$ satisfies (10) such that $n^{1/2}(\hat{\theta}_o - \theta)$ is bounded in probability and

$$\hat{\theta}_k = \hat{\theta}_{k-1} + \frac{h_f(\hat{\theta}_{k-1})}{\Sigma w_i(\hat{\theta}_{k-1})}, \qquad k = 1,\ 2,\ \ldots \tag{17}$$

Then

$$n^{1/2}(\hat{\theta}_k - \theta) \xrightarrow{D} N(0,\ V_k(s,f,g)) \tag{18}$$

where

$$\begin{aligned} V_k(s,f,g) &= \frac{I(S)}{I^2(s,g)}\left(1 - \frac{I(f,g)}{J(f,g)}\right)^{2k} \\ &+ \frac{2I(s,f)}{I(s,g)I(f,g)}\left(1 - \frac{I(f,g)}{J(f,g)}\right)^{k}\left[1 - \left(1 - \frac{I(f,g)}{J(f,g)}\right)^{2}\right] \\ &+ \frac{I(f)}{I^2(f,g)}\left[1 - \left(1 - \frac{I(f,g)}{J(f,g)}\right)^{k}\right]^2. \end{aligned} \tag{19}$$

We see that in (19) if $[1 - I(f,g)/J(f,g)]^k \longrightarrow 0$ as $k \to \infty$ then

$$\text{asyvar } n^{1/2}\hat{\theta}_k \longrightarrow I(f)/I^2(f,g) \quad \text{as } k \longrightarrow \infty, \tag{20}$$

i.e., the fully iterated WLS-rank estimate has the same asymptotic variance as the nonlinear rank estimate. The effect on the variance due to the weighted least squares method represented by $J(f,g)$, and the effect of the initial estimate, represented by $I(s,g)$ and $I(s,f)$, vanish as k approaches infinity. Further, if we are lucky enough to choose F to match G, the underlying

distribution, then $I(f,g) = I(f)$, and

$$\text{asyvar } n^{1/2}\hat{\theta}_k \doteq 1/I(f).$$

That is, the WLS-rank estimate is approximately asymptotically efficient for large k.

To show that $[1 - I(f,g)/J(f,g)]^k \to 0$ is equivalent to proving that

$$0 < I(f,g) < 2J(f,g) \; . \tag{21}$$

<u>Theorem 3</u>. Let $G \in \Omega$ and assume that ϕ_f is increasing on $(0,\infty)$, then (21) holds if either

(1) $\phi_f(G(t))/t$ is nonincreasing on $(0,\infty)$

or (2) $\phi_f(G(t))$ is concave on $(0,\infty)$.

<u>Proof</u>. First note that $I(f,g) = 2\int_0^\infty \phi_f(t)\phi_g(t)dt = 2\int_0^\infty \phi_f'(G(t))g^2(t)dt$ and $J(f,g) = 2\int_0^\infty (\phi_f(G(t))/t)g(t)dt$. Since $\phi_f(t) = -\phi_f(1-t)$, $0 \le t \le 1$ and $g(-t) = g(t)$, $-\infty < t < \infty$, condition (1) implies that on $(0,\infty)$

$$\frac{d}{dt}\left[\frac{\phi_f(G(t))}{t}\right] = \frac{\phi_f'(G(t))g(t)t - \phi_f(G(t))}{t^2} \le 0.$$

Hence

$$\int_0^\infty \phi_f'(G(t))g^2(t)dt \le \int_0^\infty \frac{\phi_f(G(t))}{t}\,g(t)dt$$
$$< 2\int_0^\infty \frac{\phi_f(G(t))}{t}\,g(t)dt.$$

Denote $\phi_f(G(t))$ by $k(t)$, condition (2) implies that on $(0,\infty)$

$$k'(t) < \frac{k(t) - k(0)}{t - 0} = \frac{k(t)}{t} \; .$$

Substitute for k(t) and simplify to obtain the result.

Corollary. Suppose F is the logistic distribution so ϕ_f is the Wilcoxon score function and suppose G is symmetric and concave on $(0,\infty)$. Then (21) holds.

The WLS-rank estimate depends on the initial estimate $\hat{\theta}_o$. In applications we would use either the sample median or the sample mean as the initial estimate. In Section 3 we discuss the stability of WLS-rank estimates using these two initial estimates.

3. Asymptotic Efficiency

We compute the asymptotic efficiency of the WLS-rank estimates relative to the maximum likelihood estimate so we can find the efficiency loss due to the weighted least squares rank estimation procedure. The values can be compared to one, the optimal value. We calculate the asymptotic efficiency for several different combinations of the initial estimate, the score generating function and the underlying distribution. In all cases the one-step WLS-rank estimate seems quite efficient in comparison with the fully iterated ones. The effect of the initial estimate wears off quickly as k increases.

- Table 1 about here -

We also calculate the asymptotic efficiencies for the WLS-rank estimator when the underlying distribution is a contaminated normal $G(x) = (1 - \varepsilon)\Phi(x) + \varepsilon\Phi(x/\sigma)$, and use the Wilcoxon scores. The table suggests that the one-step estimate is quite efficient; the efficiency converges rapidly and the effect of the initial estimate wears off quickly.

- Tables 2 and 3 about here -

4. Robustness Properties

For robustness properties we show here the stylized sensitivity curves (Andrews, et al. 1972, Section 5E) of the WLS-rank estimates.

The sample size n is taken to be 20 and the sensitivity curve is stylized by taking a pseudo sample consisting of 19 expected normal order statistics. An additional point x is then added as the 20th observation. The value of the estimator evaluated for the 20 points is denoted $T(x)$. The sensitivity curve is defined by $nT(x)$ and represents the change in T caused by adding an additional point x. Note that for the 19 centered, expected order statistics the value of T is zero. For $T = \overline{X}$ we have $nT(x) = x$, linear and unbounded, while for $T = \tilde{X}$, the median, the sensitivity curve is bounded and flat outside a neighborhood of zero. Sensitivity curves provide a finite sample analog of the influence curve discussed by Hampel (1974). Unbounded sensitivity indicates that the estimator can be unduly influenced by a small part of the data.

Figures 1-4 show the stylized sensitivity curves for the one and five step WLS-rank estimate with Wilcoxon scores and initial estimates $\bar{x}$ and $\tilde{x}$, the sample mean and median, respectively.

If we begin with the median then the WLS-rank estimate has bounded sensitivity at the first step. If we begin with the mean, with unbounded sensitivity, the WLS-rank estimate is less stable at the first step but the sensitivity becomes bounded as we take a few iterations. This nicely illustrates how the effect of the initial estimate wears off rapidly with a few iterations.

- Figures 1 - 4 about here -

It would seem preferable to use the median as a starting value; however, in more complex designs we may only have least squares starts available. Generally, two or three iterations should be sufficient with a resistant start and five or so should be sufficient with a least squares start to stabilize the WLS-rank estimate.

In Section 5F of the Princeton Robustness Study, stylized breakdown bounds are defined for estimators. For a random sample of size n, j sample points are taken to be 100,200, ..., j(100). The remaining n - j points are taken to be the n - j expected normal order statistics from a sample of size n - j. The estimator is said to break down if the resulting estimate is greater than three. Denote by m the largest j for which the estimate is less than three; m/n x 100% is then recorded. The numbers in Tables 4 and 5 are those for the WLS-rank estimates. Five iterations are used. Also included are the breakdown bounds for the mean, median, and the Hodges-Lehmann estimate so the values can readily be compared with each other.

- Tables 4 and 5 about here -

From these tables we can see that most WLS-rank estimates have larger breakdown bounds than the sample mean when the mean is used as the initial estimate. Using the median as the initial estimate the breakdown bounds are smaller than that of the median, but they are pretty close if we only take one or two iterations. This again shows the robustness of the WLS-rank estimates.

5. Extension to Regression with an Example

Adichie (1967) was the first to derive estimates of the regression coefficients in the simple linear regression model using the Hodges-Lehmann (1963) estimation procedure. The methods used by Jureckova (1971) and Jaeckel (1972) for multiple regression can be considered a generalization of the methods of Hodges and Lehmann (1963) and Adichie (1967). We shall use Jaeckel's measure of dispersion of the residuals to derive the WLS-rank estimates of the regression parameters.

Kraft and van Eeden (1972b) proposed both linearized rank and signed rank estimates for the linear model and showed that under regularity conditions, the estimates are asymptotically normal. We will apply some of their results in the discussion below.

Let Y be an n x 1 vector of observations such that

$$Y = 1\beta_o + X\beta_c + e \qquad (22)$$

where $X = [1, X_1]$ is a known n x (p + 1) matrix of full column rank, β_o is the intercept parameter and β_c is a p x 1 vector of regression parameters. Assume that the components of e are iid and each has a distribution $G \varepsilon \Omega$. Let $R(Y_i - X_i'\beta_c)$ denote the rank of $Y_i - X_i'\beta_c$ among $Y_1 - X_1'\beta_c, \ldots, Y_n - X_n'\beta_c$. Jaeckel's (1972) estimate of β_c is a value $\hat{\beta}_c$ which minimizes the convex function

$$D(\beta_c) = \Sigma a(R(Y_i - X_i'\beta_c))(Y_i - X_i'\beta_c) \qquad (23)$$

where $a(i) = \phi(i/n + 1)$ may be generated by centered versions of the score functions introduced in Section 2, namely $\phi(u) = \phi_f(u) - \overline{\phi}_f$ where $\overline{\phi}_f = \int\phi_f(u)du$. Because $\int\phi(u)du = \Sigma a(i) = 0$ we need to estimate β_o separately. This can be done as in the one-sample location problem using the residuals.

Differentiating (23) with respect to β_j, we obtain

$$\Sigma X_{ij} a(R(Y_i - X_i'\hat{\beta}_c)) = 0 \qquad j = 1, 2, \ldots, p. \tag{24}$$

Using the same technique as for the one-sample problem, equation (24) can be written in matrix form

$$X_1 \, d(\beta_o,\beta_c)X_1\beta_c = X_1 d(\beta_o,\beta_c)(Y - \beta_o 1) \tag{25}$$

where $d(\beta_o,\beta_c) = \text{diag}(W_1(\beta_o,\beta_c), W_2(\beta_o,\beta_c), \ldots, W_n(\beta_o,\beta_c))$ and

$$W_i(\beta_o,\beta_c) = \begin{cases} \dfrac{a(R(Y_i - X_i'\beta_c))}{Y_i - \beta_o - X_i'\beta_c} & Y_i - \beta_o - X_i'\beta_c \neq 0 \\ 0 & \text{Otherwise.} \end{cases} \tag{26}$$

From (25) we can define the one-step WLS-rank estimate of β_c as follows. Let $\hat{\beta}_o^{(0)}$ and $\hat{\beta}_c^{(0)}$ be initial estimates of β_o and β_c, respectively, then

$$\hat{\beta}_c^{(1)} = \hat{\beta}_c^{(0)} + [X_1'd(\hat{\beta}_o^{(0)},\hat{\beta}_1^{(0)})X_1]^{-1}X_1'd(\hat{\beta}_o^{(0)},\hat{\beta}_c^{(0)})x$$

$$a(R(Y - \hat{\beta}_o^{(0)}1 - X_1\hat{\beta}_c^{(0)})) \tag{27}$$

Denote $\psi(\beta_c) = (a(R(Y_1 - X_1'\beta_c)), \ldots, a(R(Y_n - X_n'\beta_c)))'$, then

$$\hat{\beta}_c^{(1)} = \hat{\beta}_c^{(0)} + [X_1'd(\hat{\beta}_o^{(0)},\hat{\beta}_c^{(0)})X_1]^{-1}X_1'\psi(\hat{\beta}_c^{(0)}) \tag{28}$$

which is similar to Bickel's (1975) one-step Huber M-estimate of type I and Beaton and Tukey's (1974) weighted least squares M-estimate. We shall call it the one-step WLS-rank estimate of type I. $\hat{\beta}_o^{(1)}$ is obtained using the one-sample procedure. We would generally use least squares estimates to start the iteration. Our estimate is more complicated than the ones proposed by McKean and Hettmansperger (1978) and Kraft and van Eeden (1972b).

Nevertheless, we do not need to estimate a scale parameter as they do. To develop the asymptotic distribution, we rewrite the model using the centered design matrix,

$$Y = 1\beta_o + X_1\beta_c + e = \beta_o^* + X_{1c}\beta_c + e$$

where $X_{1c} = X_1 - \overline{X}_1$.

Following the approach of Bickel (1975), define the one-step WLS-rank estimate of type II, $\hat{\beta}_c^{*(1)}$ as

$$\hat{\beta}_c^{*(1)} = \hat{\beta}_c^{(0)} + \frac{1}{\hat{J}(f,g)} (X_{1c}'X_{1c})^{-1}X_{1c}\psi(\hat{\beta}_c^{*(0)}) \tag{29}$$

where $\hat{J}(f,g)$ is a consistent estimate of J(f g). The asymptotic normality of $\hat{\beta}_c^{*(1)}$ under regularity conditions then can be established following the proof of Kraft and van Eeden (1972b) (see Cheng (1979)). We state the result in the following theorem.

Theorem 4. Under the regularity conditions of Kraft and van Eeden (1972b), $n^{1/2}(\hat{\beta}_c^{*(1)} - \beta_c)$ has, asymptotically, a multivariate normal distribution with mean vector 0 and covariance matrix given by

$$\left[\frac{I(S)}{I(s,g)^2}\left(1 - \frac{I(f,g)}{J(f,g)}\right)^2 + \frac{2I(s,f)}{I(s,g)I(f,g)}\left(1 - \frac{I(f,g)}{J(f,g)}\right) + \frac{I(f)}{J(f,g)^2}\right] \Sigma_{1c}^{-1}$$

where Σ_{1c} is positive definite and $n^{-1}X_{1c}\ X_{1c} \to \Sigma_{1c}$

The asymptotic distribution of the WLS-rank estimate of type I is the same as that for type II. The proof requires further regularity conditions on the weights and follows along the lines of Bickel (1975); see Cheng (1979) for details. Hence the asymptotic covariance matrix contains the

factor (15) in the one-step case. It can be shown in the same way that factor (19) appears for the k-step case.

As an example we consider the stack loss data analyzed by Daniel and Wood (1971) in their Chapter 5. The example contains 21 observations and 3 parameters. Daniel and Wood studied the problem extensively using least squares and found 4 outliers. They fitted the model after removing the outliers. Using a robust regression procedure, Andrews (1974) obtained a suitable fit without deleting the outliers as did Hettmansperger and McKean (1977) using rank estimates.

An APL program was written to calculate the WLS-rank estimates. The initial estimates were least squares estimates, and negative weights were set zero. In order to check the convergence of the iteration procedure, thirty iterations were used. In the following table we show the k-step WLS-rank estimates using sign and Wilcoxon scores. R-estimates obtained by Hettmansperger and McKean (1977) were also included. The estimates were quite close. This should be the case since the same dispersion function was used, only the techniques used to obtain the estimates were different.

Hence the weighted least squares approach achieves an acceptable solution without searching a convex surface which may be costly or requiring the estimation of an auxilliary scale parameter.

- Table 6 about here -

REFERENCES

1. Andrews, D. F. (1974). "A Robust Method for Multiple Linear Regression", Technometrics, 16, 523-31.
2. Andrews, D. F., Bickel, P. J., Hampel, R. F., Huber, P. J., Rogers, W. H. and Tukey, J. W. (1972). Robust Estimates of Location: Survey and Advances. New Jersey: Princeton University Press.
3. Bauer, D. F. (1972). "Constructing Confidence Sets Using Rank Statistics", J. Amer. Statist. Assoc., 38, 894-904.
4. Beaton, A. E. and Tukey, J. W. (1974). "The Fitting of Power Series, Meaning Polynomials, Illustrated on Band-Spectroscopic Data", Technometrics, 16, 147-185.
5. Bickel, P. J. (1975). "One Step Huber Estimates in the Linear Model", J. Amer. Statist. Assoc., 70, 428-434.
6. Cheng, K. (1979). On Weighted Least Squares Estimates. Unpublished Ph.D. Thesis, Penn State University.
7. Daniel, C. and Wood, F. S. (1971). Fitting Equations to Data. New York: John Wiley and Sons.
8. Hajek, J. and Sidak, Z. (1967). Theory of Rank Statistics. New York: Academic Press.
9. Hampel, F. R. (1974). The Influence Curve and Its Role in Robust Estimation", J. Amer. Statist. Assoc., 69, 383-93.
10. Hettmansperger, T. P. and McKean, J. W. (1977). "A Robust Alternative Based on Ranks to Least Squares in Analyzing Linear Models", Technometrics, 19, 275-284.

11. Hettmansperger, T. P. and Utts, J. M. (1977). "Robustness Properties for a Simple Class of Rank Estimates", Communications in Statistics - Theory and Methods, A6(9), 855-868.

12. Hodges, J. L., Jr. and Lehmann, E. L. (1963). "Estimation of Location Based on Rank Tests", Ann. Math. Statist. 34, 598-611.

13. Holland, P. W. and Welsch, R. E. (1977). "Robust Regression Using Iteratively Reweighted Least Squares", Communications in Statistics - Theory and Methods, A6(9), 813-827.

14. Huber, P. J. (1981). Robust Statistics. New York: John Wiley and Sons.

15. Jaeckel, L. A. (1972). "Estimating Regression Coefficients by Minimizing the Dispersion of the Residuals", Ann. Math. Statist. 43, 1449-58.

16. Jureckova, J. (1969). "Asymptotic Linearity of a Rank Statistic in Regression Parameter", Ann. Math. Statist. 40, 1889-1900.

17. Jureckova, J. (1971). "Nonparametric Estimate of Regression Coefficients", Ann. Math. Statist. 42, 1328-38.

18. Kraft, C. H. and van Eeden, C. (1970). "Efficient Linearized Estimates Based on Ranks", Nonparametric Techniques in Statistical Inference, ed. M. L. Puri, Cambridge Univ. Press, 267-273.

19. Kraft, C. H. and van Eeden, C. (1972a). "Relative Efficiencies of Quick and Efficient Methods of Computing Estimates from Rank Tests", J. Amer. Statist. Assoc., 67, 199-202.

20. Kraft, C. H. and van Eeden, C. (1972b). "Linearized Rank Estimates and Signed-Rank Estimates for the General Linear Hypothesis", Ann. Math. Statist. 43, 42-57.

21. McKean, J. W. and Hettmansperger, T. P. (1978). "A Robust Analysis of the General Linear Model Based on One-Step R-Estimates", Biometrika, 65, 571-579.

22. Utts, J. (1978). A Robust Class of Tests and Estimates for Multivariate Location Based on Ranks. Unpublished Ph.D. Thesis, Penn State University.

23. van Eeden, C. (1972). "A Analogue, for Signed Rank Statistics, of Jureckova's Asymptotic Linearity Theorem for Rank Statistics", Ann. Math. Statist., 43, 791-802.

Table 1. Asymptotic efficiencies of k-step WLS-rank estimates.

	k						
	1	2	3	4	5	10	30
Sign scores, $\hat{\theta}_o = \bar{X}$							
Normal	.937	.847	.776	.728	.697	.644	.637
Logistic	.945	.957	.955	.943	.928	.844	.755
D.E.	.592	.677	.753	.815	.865	.976	1.0
Wilcoxon scores, $\hat{\theta}_o = \bar{X}$							
Normal	.971	.959	.956	.955	.955	.955	.955
Logistic	.995	1.0	1.0	1.0	1.0	1.0	1.0
D.E.	.689	.735	.747	.75	.75	.751	.751
van der Waerden scores, $\hat{\theta}_o = \bar{X}$							
Normal	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Logistic	.956	.955	.955	.955	.955	.955	.955
D.E.	.628	.636	.637	.637	.637	.637	.637
Wilcoxon scores, $\hat{\theta}_o = \tilde{X}$							
Normal	.919	.950	954	.955	.955	.955	.955
Logistic	.984	.999	1.0	1.0	1.0	1.0	1.0
D.E.	.853	.780	.759	.753	.752	.751	.751
van der Waerden scores, $\hat{\theta}_o = \tilde{X}$							
Normal	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Logistic	.947	.955	.955	.955	.955	.955	.955
D.E.	.667	.639	.637	.637	.637	.637	.637

Table 2. Asymptotic efficiencies of k-step WLS-rank estimates for contaimınated normal distributions with Wilcoxon Scores.

		k						
		1	2	3	4	5	10	30
$\hat{\theta}_o = \overline{X}$								
	$\sigma = 2$	.989	.993	.993	.993	.993	.993	.993
$\varepsilon = .2$	3	.892	.923	.927	.927	.927	.927	.927
	4	.791	.848	.856	.857	.857	.857	.857
	$\sigma = 2$	.987	.987	.986	.986	.986	.986	.986
$\varepsilon = .1$	3	.941	.956	.958	.958	.958	.958	.958
	4	.881	.916	.920	.920	.920	.920	.920
	$\sigma = 2$	.983	.979	.979	.979	.979	.979	.979
$\varepsilon = .05$	3	.964	.969	.970	.970	.970	.970	.970
	4	.931	.947	.949	.949	.949	.949	.949

Table 3. Asymptotic efficiencies of k-step WLS-rank estimates for contaminated normal distributions with Wilcoxon Scores.

		k						
		1	2	3	4	5	10	30
$\hat{\theta}_o = \tilde{X}$								
	$\sigma = 2$	.984	.993	.993	.993	.993	.993	.993
$\varepsilon = .2$	3	.928	.928	.927	.927	.927	.927	.927
	4	.864	.858	.857	.857	.857	.857	.857
	$\sigma = 2$	.973	.985	.986	.986	.986	.986	.986
$\varepsilon = .1$	3	.950	.958	.958	.958	.958	.958	.958
	4	.915	.920	.920	.920	.920	.920	.920
	$\sigma = 2$	.963	.977	.979	.979	.979	.979	.979
$\varepsilon = .05$	3	.956	.969	.970	.970	.970	.970	.970
	4	.938	.948	.949	.949	.949	.949	.949

Table 4. The breakdown bounds for k-step WLS-rank estimates (initial estimate: $\overline{X}$, scores: Wilcoxon).

		Sample Size			
		5	10	20	40
	k=0* (mean)	0	0	0	2.5
	1	0	10	5	7.5
	2	0	10	10	10
Number of Iterations	3	20	10	15	15
	4	20	20	15	17.5
	5	20	20	20	17.5
	∞* (Hodges-Lehmann estimate)	20	20	25	27.5

*Entries in these rows are from the Princeton Robustness Study (Andrews, et al., 1972, Section 5F).

Table 5. The breakdown bounds for k-step WLS-rank estimates (initial estimate: $\tilde{X}$, scores: Wilcoxon).

	Sample Size			
	5	10	20	40
$k=0^*$ (median)	40	40	45	47.5
1	40	40	45	42.5
2	40	40	40	37.5
3	20	30	35	37.5
4	20	30	35	37.5
5	20	30	30	32.5
∞^* (Hodges-Lehmann estimate)	20	20	25	27.5

*Entries in these rows are from the Princeton Robustness Study (Andrews, et al., 1972, Section 5F).

Table 6. Estimates of regression coefficients.

Method		α	β_1	β_2	β_3
Least squares		-39.9	.72	1.30	-.15
Least squares w/o outliers		-37.6	.80	.58	-.07
Andrews		-37.2	.82	.52	-.07
Hettmansperger and McKean:	Sign	-39.7	.83	.58	-.06
	Wilcoxon	-39.95	.80	.90	-.11
WLS-rank, Sign,	k = 1	-40.01	.805	.903	-.114
	2	-40.02	.839	.718	-.096
	3	-40.01	.837	.649	-.078
	4	-40.01	.837	.603	-.067
	5	-40.00	.831	.588	-.060
	10	-40.00	.833	.567	-.057
	30	-40.00	.833	.568	-.057
Wilcoxon,	k = 1	-40.29	.804	1.006	-.138
	2	-40.36	.804	.955	-.126
	3	-40.50	.809	.913	-.118
	4	-40.50	.804	.907	-.113
	5	-40.50	.799	.930	-.115
	10	-40.50	.810	.904	-.116
	30	-40.50	.813	.893	-.116

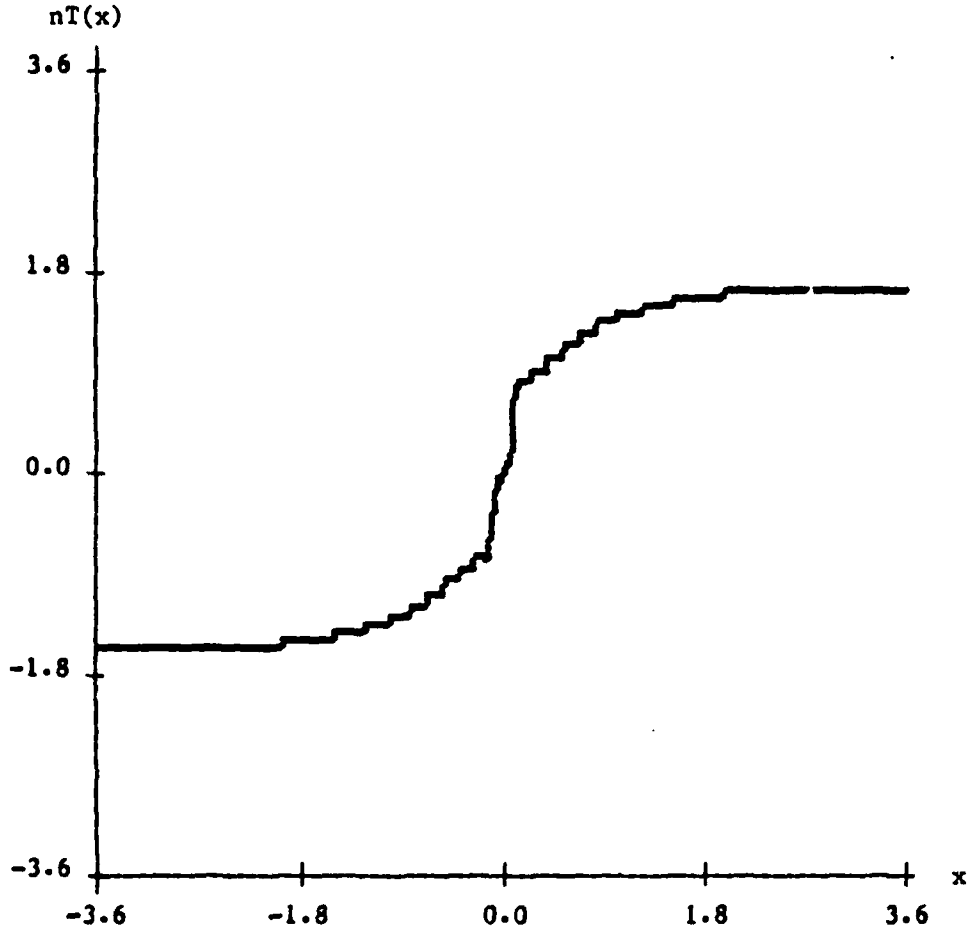


Figure 1. Stylized sensitivity curve for 1-step WLS-rank estimate (initial estimate: $\tilde{x}$, scores: Wilcoxon).

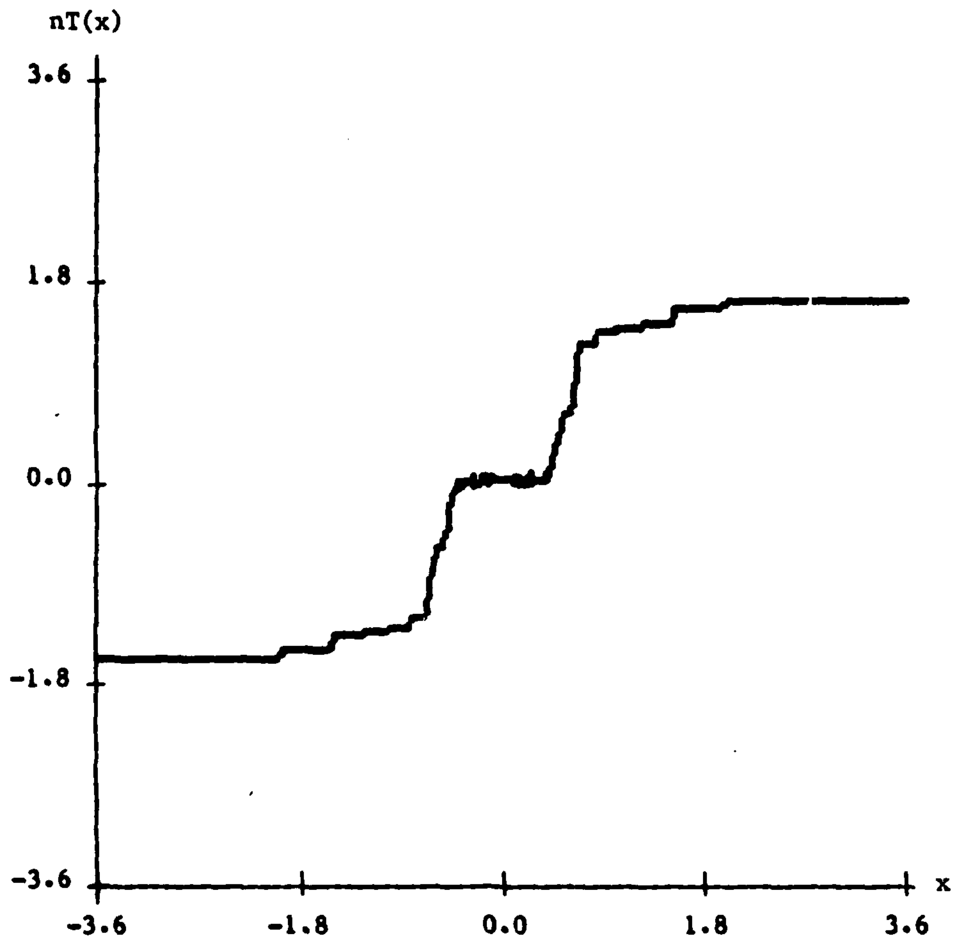


Figure 2. Stylized sensitivity curve for 5-step WLS-rank estimate (initial estimate: $\tilde{x}$, scores: Wilcoxon).

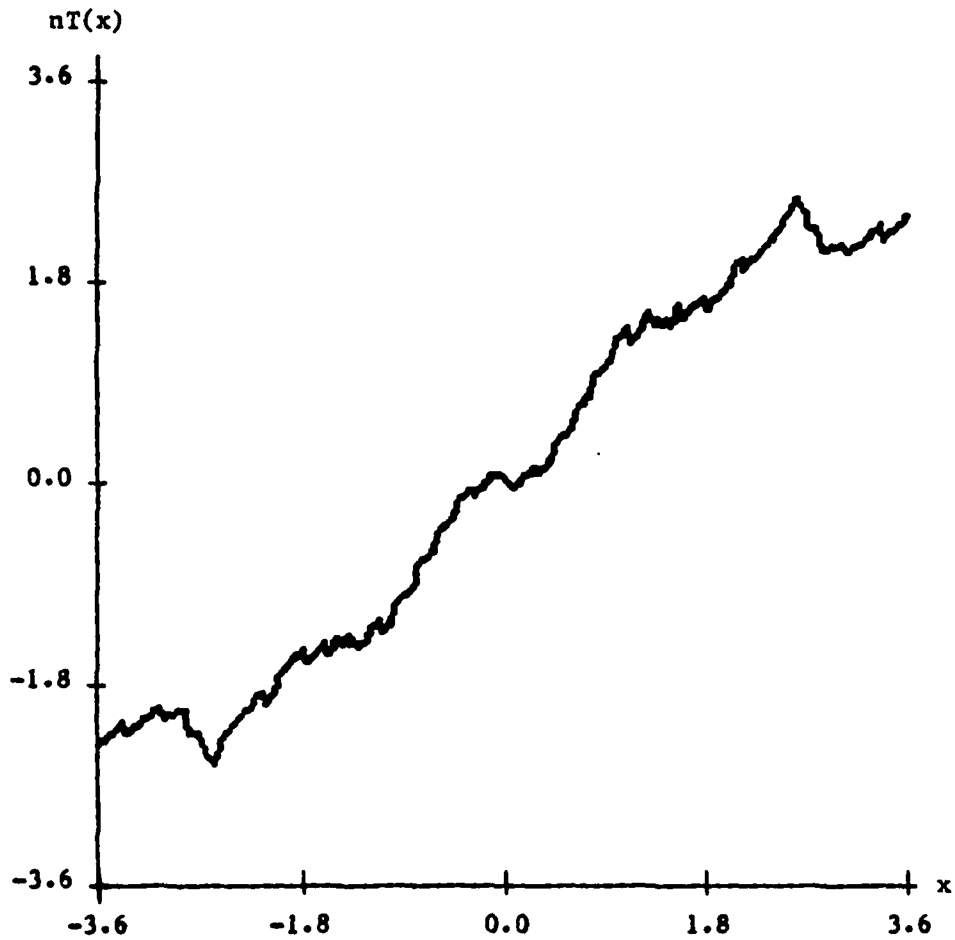


Figure 3. Stylized sensitivity curve for 1-step WLS-rank estimate (initial estimate: $\bar{x}$, scores: Wilcoxon).

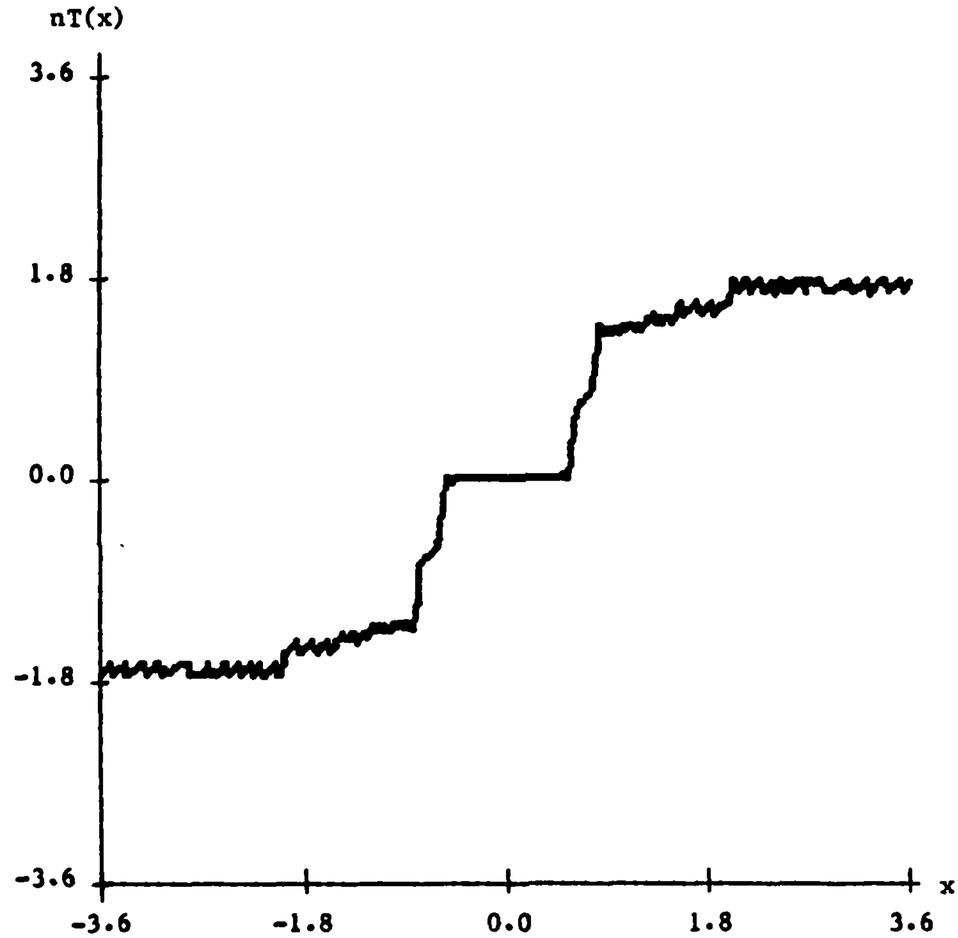


Figure 4. Stylized sensitivity curve for 5-step WLS-rank estimate (initial estimate: $\bar{x}$, scores: Wilcoxon).

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Penn State Tech. Rpt. 38	2. GOVT ACCESSION NO. AD-A102759	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Weighted Least Squares Rank Estimates		5. TYPE OF REPORT & PERIOD COVERED 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Kuo-Sheuon Cheng, Tulane University Thomas P. Hettmansperger, Penn State University		8. CONTRACT OR GRANT NUMBER(s) N00014-80-C-0741
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Penn State University University Park, PA 16802		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 042-446
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics and Probability Program Code 436 Arlington, VA 22217		12. REPORT DATE July 1981 13. NUMBER OF PAGES 31
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Robust estimates, robust confidence intervals, nonparametric tests

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

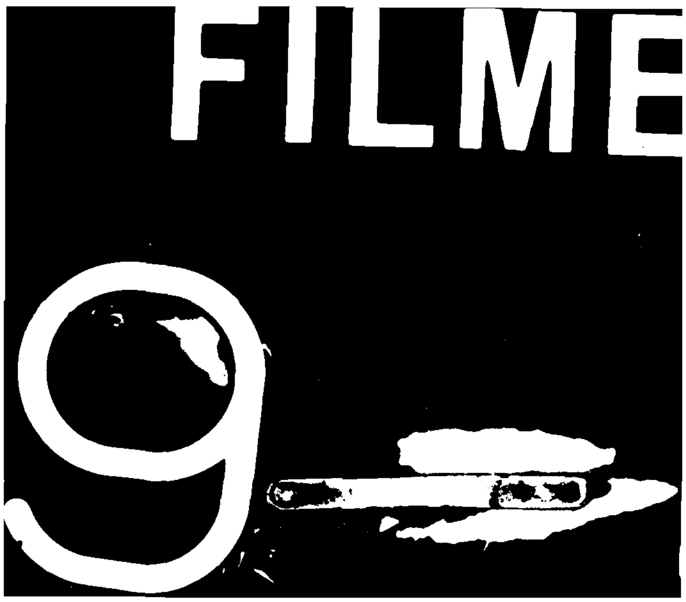
In this paper rank estimates, called WLS rank estimates and computed using iteratively reweighted least squares, are studied. They do not require the estimation of auxilliary scale or slope parameters nor do they require numerical search techniques to minimize a convex surface. The price is a small asymptotic efficiency loss. In the location model, beginning with a resistant starting value such as the median, the WLS rank estimates have good robustness and computational properties. The WLS rank estimate is also extended to the regression model and an example is given.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)


FILME
9